DOCUMENT RESUME

ED 135 822 TM 006 046

AUTHOR Conard, C. J.; And Others
TITLE Self-Grading versus External Proctoring: A Counterbalanced Comparison.
PUB DATE [76]
NOTE 10p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *College Students; Comparative Analysis; Grades (Scholastic); *Grading; Higher Education; *Individualized Instruction; *Proctoring; *Self Evaluation; Student Attitudes; *Student Evaluation

ABSTRACT

This study compared external proctoring and student self-grading in a personalized child development course. The experiment used a counterbalanced experimental design and two traditional control groups. Survey data and objective preferences indicated that students preferred self-grading to proctor-grading. However, students reported that proctor-grading prepared them better for major review exams. Although this belief was not supported by hour exam data from the two counterbalanced groups, results from the traditional control groups indicated that self-grading produced performance that was 10 percentage points lower than proctor-grading. These results are discussed in terms of the use of self-grading procedures in self-paced, individualized courses. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED135822

Self-Grading versus External Proctoring:

A Counterbalanced Comparison

C. J. Conard, Robert E. Spencer and George Semb

Department of Human Development

University of Kansas

Lawrence, Kansas 66045

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Abstract

This study compared external proctoring and student self-grading in a personalized child development course. The experiment used a counterbalanced experimental design and two traditional control groups. Survey data and objective preferences indicated that students preferred self-grading to proctor-grading. However, students reported that proctor-grading prepared them better for major review exams. Although this belief was not supported by hour exam data from the two counterbalanced groups, results from the traditional control groups indicated that self-grading produced performance that was 10 percentage points lower than proctor-grading. These results are discussed in terms of the use of self-grading procedures in self-paced, individualized courses.

Introduction

Since the inception of personalized instruction in 1968 (Keller, 1968), much research has been conducted to analyze the effects and efficiency of its various components. Component analysis has, for example, affirmed the importance of study questions (Semb, Hopkins, & Hursh, 1973), unit assignments (Semb, 1974a), high mastery criteria (Johnston & O'Neill, 1973) and external proctors (Farmer, Lachter, Blaustein, & Cole, 1972). While these components have been validated as effective and critical determiners of student performance, one of them, the use of external proctors, appears to have additional benefits. Not only do proctors produce high exam performances in their students (Farmer, et al., 1972), they also serve to facilitate the ease with which other components of personalized instruction are implemented. For example, instructors who have their course divided into small units of material frequently use proctors for the frequent quizzing, grading, and feedback that this component requires. Thus, many instructor related duties are handled by the proctor.

Despite these benefits, some instructors may be unable or unwilling to use external proctors. First, external proctors must be selected, trained, and monitored to insure that they grade quizzes accurately (Semb, 1975a).

TM006 046

Second, few instructors have the financial support to pay proctors for their services. One alternative has been to offer proctors academic credit. However, problems may arise if the educational setting does not permit proctors to receive course credit for this task. While the proctoring experience may provide an excellent opportunity for students to interact with their peers and develop valuable social skills, administrators may argue that the instruction belongs in the hands of the instructor and that proctors are not considered faculty. Some administrators may also argue that the proctors do not profit academically from this experience. Without money or credit to offer proctors, an instructor may decide not to adopt a personalized format. Even instructors who have access to course credit or financial support for proctors still have the burdensome task of recruiting applicants, selecting those best qualified, training appropriate proctor behaviors, and staffing.

At least two alternative systems -- internal proctoring, which uses currently enrolled students to evaluate peers quizzes (Gaynor & Wolking, 1974; Johnson & Sulzer-Azaroff, 1974), and self-proctoring, which utilizes students to evaluate their own quizzes (Blackburn, Semb, & Hopkins, 1975) -- are currently available to use in place of external proctors. One possible problem with internal proctoring is that students may be reluctant to have their performance evaluated by classmates or to act as peer-graders. On the other hand, students may be receptive to a procedure whereby they evaluate their own performance (i.e., a system of self-proctoring or self-grading). As suggested by Gagne (1965) "...the student must be progressively weaned from dependence on the teachers or other agents external to himself." Blackburn, Semb, and Hopkins (1975) recently demonstrated that self-grading is effective in maintaining high levels of academic performance on review tests and a final examination. A follow-up study by Blackburn, Semb, and Hopkins (1974) demonstrated that the number of proctors could be reduced by 50% without any loss in classroom efficiency or student performance. Their results suggest that self-grading is a viable alternative for instructors who wish to use a personalized format. However, no data have been collected which analyze student preference for the self-grading procedure. The present study compares self-grading to external proctoring in a self-paced, personalized child development course. The dependent measures are student performance on major exams and student preference for the two systems.

Method

Subjects, Setting, and Course Personnel

Seventy-two students enrolled in two sections of an introductory child development course served as subjects. Twelve students withdrew from the course, leaving sixty students who participated in the study. Students were randomly assigned to one of four groups in each section. The two sections operated at the same time in two adjacent rooms. Each section was staffed by one graduate teaching assistant, four external proctors, and one administrative assistant. Each proctor was responsible for nine students.

General Course Format and Procedures

The course content was divided into three major parts; each part was further subdivided into five units. Each unit consisted of approximately one chapter (30-40 pages) from the texts (Lefrancois, 1973; Semb, 1975b) and an accompanying chapter from the study guides (Semb, 1974b; Semb, 1975b).

The course was self-paced to the extent that students could work as fast as they wanted and instructor paced in that students were required to maintain a minimum rate of progress or drop the course. The semester lasted approximately 14 weeks or 40 class days.

In order for students to complete the course, they were required to complete 15 unit quizzes, three review exams, and a final. Unit quizzes consisted of six questions sampled from a pool of 20-30 items. Three forms of each unit quiz were constructed by randomly assigning questions to each form. The remaining unselected questions were used to construct 15-item review exams (three items from each unit). Review exam items were randomly assigned to forms such that each student received a different exam. All unit quizzes and review exams were distributed by an "administrative assistant." That is, when the students were ready to take a quiz or exam, they reported to the assistant who gave them the appropriate test. After a student had completed all quizzes and hour exams, a comprehensive final was given. The final consisted of 90 true-false items (six items from each unit). At this time the student also completed a short evaluation which was attached to the final.

All review exams were graded outside of class by an external grader. For experimental purposes, the same grader was used for all review exams to insure grading consistency. Review exams could be retaken once; the highest of the two scores counted. Alternate forms of the review exam were generated by randomly selecting 15 items from the hour exam item pool.

Unit quizzes were evaluated according to one of the grading conditions described below.

Proctor grading. Students gave the quiz to the proctor who then graded it according to an answer key. Items were graded as worth 0, 1, or 2 points. If performance was less than 10 out of the 12 possible, the quiz was filed and a retake was required. If the performance was 10 or higher, the student could discuss the errors and then make written corrections to bring the quiz to a 100% mastery level. After the quiz was marked as complete and correct, the proctor would ask the student to explain two concepts from the unit. Concepts were randomly preselected for each unit but students were not informed which ones had been selected until they successfully completed the unit quiz. When a satisfactory verbal explanation of the concepts was given, the student was considered "passed" and allowed to continue on to the next unit.

Self-Grading. Students took the completed quiz to the administrative assistant who issued an answer key. Students then evaluated their own responses by comparing them to the answer key. The same criteria were used for self-grading. Less than 10 out of 12 required a retake, whereas 10 or better could be remediated to the 100% mastery level. After writing corrected answers, students gave the quiz to their proctor who then conducted the same discussion over concepts as described in the proctor grading condition.

Students final grades were determined by their performance on the unit quizzes (40%), review exams (40%), and the final exam (20%).

Experimental Design

The experiment used a counterbalanced reversal design (Semb, 1976) with a forced choice component as illustrated in Table 1. Groups 1 and 2 provided a counterbalanced comparison between external proctoring and self-grading. Groups 3 and 4 served as traditional control groups that allowed the assessment of the effects of continued exposure to an experimental condition. All groups had a choice between the two conditions during the third part of the course. The choice was available for each unit of Part 3.

Table 1

The Experimental Design

	Course Parts		
	1	2	3
Group 1	External proctoring	Self-grading	Choice
Group 2	Self-grading	External proctoring	Choice
Group 3	Self-grading	Self-grading*	Choice
Group 4	External proctoring	External proctoring*	Choice

* To expose students in Groups 3 and 4 to the alternate procedure prior to the choice condition, they were required to complete the last unit of Part 2 under the alternate condition.

Group One. Students progressed through Part 1 of the course under the proctor-grading condition. During Part 2, the self-grading procedure was implemented.

Group Two. Students operated under self-grading in Part 1 and then switched to proctor-grading for Part 2.

Group Three. Students had their unit quizzes evaluated under the self-grading procedure for both Parts 1 and 2, with the exception of the last unit in Part 2 when they were exposed to proctor-grading.

Group Four. Students progressed through the first two parts of the course under the proctor grading condition, except for the last unit in Part 2 in which they were exposed to the self-grading procedure.

Evaluation

A short evaluation was attached to the final exams. Students were asked to respond to the following three questions:

(1)	Which procedure did you like best?	self-grade	proctor-grade
(2)	Which procedure do you feel helped you best prepare for review exams?	self-grade	proctor-grade
(3)	If you took another PSI course, which procedure would you want to use?	self-grade	proctor-grade

Reliability measures

Proctors regraded a total of 397 self-graded quizzes to check the accuracy with which they had been graded. Agreements were defined as proctor-student combinations of 2-2, 1-1, or 0-0 points; disagreements were defined as any discrepancy between the student and the proctor (i.e., 2-1, 2-0, 1-2, 1-0, 0-1, 0-2). Of the 2382 items regraded, there were 2286 agreements and 96 disagreements. Reliability, calculated by dividing the number of agreements by the number of agreements plus disagreements, was 0.960.

To check proctor-grading accuracy, a teaching assistant regraded three quizzes (one from each part) for each of the ten proctors. Of the 360 items regraded, there were 336 agreements and 24 disagreements. Reliability, calculated as described above, was 0.933.

Finally, a teaching assistant also regraded two hour exams from each condition for each part of the course for each of the four groups. Of the 360 items regraded (24 hour exams), there were 285 agreements and 75 disagreements. Reliability was 0.792.

Results and Discussion

Student Review Exam Performance

Performance on the hour exams must be interpreted cautiously. Grading reliability was less than 80% which indicates that grading was not as consistent as it has been in previous research in the same course (Spencer, Conyers, Sanchez-Sosa, & Semb, 1975; Semb, Spencer, & Phillips, 1976). However, there were no consistent grader biases, which suggest that errors in reliability checks were randomly distributed.

Combining results from Groups 1 and 2 (the counterbalanced groups), proctor-grading produced a mean performance of 80.7% correct on first attempt review exams, as compared with 81.2% for self-grading. If one takes retake exams into account, proctor-grading produced a mean of 82.0% as compared to 87.0% for self-grading. Thus, it would appear as if proctor-grading and self-grading produce comparable results, but that students in self-grading have a slight tendency to improve their scores when retakes are available.

Hour exam performance from the two traditional control groups (Groups 3 and 4) for Parts 1 and 2 show a slightly different pattern. Proctor-grading produced a mean performance of 80.3% on first attempt quizzes as compared with 72.9% for self-grading. Taking retake exams into account, proctor grading produced a mean of 84.3% as compared with 74.7% for self-grading, a difference of nearly 10 percentage points. Thus, it would appear as if the effects of prolonged exposure to self-grading are somewhat deliterious when compared with proctor-grading.

Student Preference - Survey

The results of the survey which accompanied the final exam were analyzed only for the two groups (1 and 2) who experienced the procedures for an entire part. Due to administrative errors, several students in the two traditional control groups (3 and 4) were not exposed to the alternate condition. Thus, their survey data are not included in the present analysis. The percentage of students who selected the self-grading procedure for each of the evaluation questions is shown in Table 2.

Table 2

Survey Results: Preference for self-grading

	Questions		
Group	Best-liked	Review Exam Preparation	Future Choice
1 (Proctor-Self-Choice)	66.5%	41.6%	58.3%
2 (Self-Proctor-Choice)	56.2%	12.5%	43.7%

Results of the survey indicate that both groups liked the self-grading procedure better than proctor-grading. Group 1, which experienced self-grading

last, indicated a higher preference for self-grading than Group 2, which experienced self-grading first. This indicates that the order of experimental conditions may have affected students' written preferences.

Both groups believed that proctor-grading prepared them better for review exams than self-grading. Furthermore, the group that experienced proctor grading last (Group 2) was almost unanimous (87.5%) in their view of proctor-grading prepared them better

Finally, Group 1 indicated a slight future preference for self-grading, while Group 2 indicated a slight future preference for proctor-grading. Although these results are at best equivocal, it is interesting to note that students who experienced self-grading last showed a slight preference for that procedure in the future, whereas the group that experienced proctor-grading last were more favorably disposed toward the future prospects of proctor-grading. Again, the order of experimental conditions may have had an effect.

Objective Choice Data

Although students' pencil and paper preferences are interesting, they may not be as convincing as the actual choices students make. Choice data (Part 3) were analyzed for the counterbalanced groups (1 and 2) to determine if the order of experimental conditions affected preference. Choice data for the traditional control groups (3 and 4) were analyzed to determine the effects of continued exposure to a procedure.

Students from Group 1 (Proctor-Self-Choice) chose to self-grade 45 unit quizzes (64%) and to have 25 (36%) proctor-graded. Students from Group 2 (Self-Proctor-Choice) chose to self-grade 46 quizzes (55%) and to have 38 (45%) graded by a proctor. Thus, it would appear as if students prefer self-grading, a finding similar to that found on the survey, regardless of the order to which they were exposed to experimental conditions. However, the effect was smaller for the group exposed to proctoring last, suggesting that order may have a slight effect.

Results from the traditional control groups must be interpreted with caution. Due to administrative errors, six students in Group 3 (Self-Self-Choice) were not exposed to the alternate procedures at Part 2, Unit 5, but at Part 3, Unit 1, whereas two students in Group 4 (Proctor-Proctor-Choice) were not exposed at Part 2, Unit 5, but at Part 3, Unit 1. Also, two students from both Groups 3 and 4 were never exposed to the alternate procedures and thus were eliminated from the data.

Students from Group 3 (Self-Self-Choice) selected self-grading for 51 of the 64 units for which it was available (79.7%). By even a greater margin, Group 4 (Proctor-Proctor- Choice) also selected self-grading (46 of the 53 units during which it was available, or 86.8%). Both of these groups showed a strong objective preference for self-grading, a result that cannot be explained by either a novelty effect or an order effect. Perhaps self-grading is a popular procedure, one that perseveres despite other, extraneous factors.

Summary

Overall, survey data and objective preferences indicate that students prefer self-grading as compared to proctor-grading. However, this preference must be tempered by the fact that the order of experimental conditions may have attenuated the effect. Furthermore, students indicated that they believed that proctor-grading prepared them better for review exams. Although this finding was not substantiated by hour exam performance data from the counterbalanced groups (1 and 2), results from the traditional control groups (3 and 4) indicate that self-grading produces performance substantially inferior (10 percentage points) to proctor-grading. These results are similar to those reported by Spencer and Semb (1976) in which students preferred the easier of two grading conditions, but performed best under the one which was the most stringent. Nevertheless, it would appear as if student self-grading, with appropriate quality control mechanisms, may be a cost-effective alternative to the use of external student proctors in self-paced, individualized courses. The use of students as their own evaluation agents deserves further experimental investigation.

References

Blackburn, T., Semb, G., Hopkins, B.L. An analysis of self-grading procedures in a course taught by personalized instruction. American Psychological Association, New Orleans, August, 1974.

Blackburn, T., Semb, G., Hopkins, B.L. The comparative effects of self-grading on classroom efficiency and student performance in a personalized instruction course. In J. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill.: Charles C. Thomas, 1975. Pp. 250-268.

Farmer, J., Lachter, G.P., Blaustein, J.J. & Cole, B.K. The role of proctoring in personalized instruction. Journal of Applied Behavior Analysis, 1972, 5, 401-404.

Gagne, R.M. The conditions of learning. New York: Holt, Rinehard, and Winston, 1965.

Gaynor, J.F. & Wolking, W.D. The effectiveness of currently enrolled student proctors in an undergraduate special education course. Journal of Applied Behavior Analysis, 1974, 7, 263-269.

Keller, F.S. "Goodbye, teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Johnson, K.R. & Sulzer-Azaroff, B. The effects of different proctoring systems on student examination performance and preference. In J. Johnston (Ed.), Behavior research and technology in higher education. Springfield, Ill.: Charles C. Thomas, 1975.

Johnston, J.M. & O'Neill, G. The analysis of performance criteria defining course grades as a determinant of college student academic performance. Journal of Applied Behavior Analysis, 1973, 6, 261-268.

Lefrancois, G.F. Of children. Belmont, Calif.: Wadsworth, 1973.

Semb, G. The effects of mastery criteria and assignment length on college student test performance. Journal of Applied Behavior Analysis, 1974, 1, 61-69, (a).

Semb, G. Of children: A study guide. Belmont, Calif.: Wadsworth, 1974 (b).

Semb, G. Behavior analysis: A practical and empirical approach to child development. Division of Continuing Education, University of Kansas, Lawrence, Kansas, 1975 (b).

Semb, G. Proctor selection, training and quality control in personalized instruction. In J. Johnston (Ed.), Research and technology in college and university instruction. Gainesville: Department of Psychology, University of Florida, 1975. Pp. 139-150, (a).

Semb, G. Building an empirical base for instruction. Journal of Personalized Instruction, 1976, 1, 11-22.

Semb, G., Hopkins, B.L., & Hursh, D.E. The effects of study questions and grades on student test performance in a college course. Journal of Applied Behavior Analysis, 1973, 6, 631-642.

Semb, G., Spencer, R.E., & Phillips, T.W. The use of review units in a personalized university course. In B.A. Green (Ed.), Personalized instruction in higher education. Washington, D.C.: Center for Personalized Instruction, Georgetown University, 1976. Pp. 140-145.

Spencer, R., Conyers, D., Sanchez-Sosa, J.J., & Semb, G. An experimental comparison of two forms of personalized instruction, a discussion proc-dure and an independent study procedure. In R. Ruskin & S. Bono (Eds.), Personalized instruction in higher education. Washington, D.C.: Center for Personalized Instruction, Georgetown University, 1975. Pp. 11-20.

Spencer, R. & Semb, G. An analysis of the effects of study questions and unit quizzes under two different grading conditions. In J.G. Sherman (Ed.), Personalized instruction in higher education III. Washington, D.C.: Center for Personalized Instruction, Georgetown University, in press.